Forecasting for Impact

Daily Demand Prediction at Walmart

Forecasting Challenge at Walmart

Situation

Walmart operates at a massive scale - forecasting daily demand for over **3,000 products across 10 stores**

Challenge

Existing models struggle with:

- **Volatile demand** across departments,
- **Sparse data** at item-store level,
- **High cost of errors**: overstock ties up capital, stockouts lose sales

Key Question

How can we develop a **scalable and accurate demand forecasting system** to guide better stocking decisions?

Data Landscape

Sales Data

6 years of daily item-store sales
3,049 products x 10 stores (Jan 2011 - Jun 2016)

Calendar Events

Holidays, SNAP days, weekends

Prices

Weekly item-store pricing

Hierarchical Structure

State

↓

Store

↓

Product Category

↓

Department

↓

Item

Model Evolution – A Scalable Forecasting Pipeline

Model Group	Business Role	Observations
LightGBM	Fast baseline model	Feature-rich but less accurate
LSTM	Memory-based model with event-aware forecasting	Better than LightGBM, struggled with overfitting
GRU	Efficient temporal model with best accuracy-to-speed ratio	**Best performer**, lean architecture, stable results
Seq2Seq	One-shot 28 day forecasting	Fastest to train, underperformed when compared to GRU

Takeaway: Advanced models excel - GRU balances accuracy and speed best, while LightGBM helps interpret but lacks performance

Features that Drive Sales

Best Model

Historical Sales Pattern

Captures recurring demand cycles and weekly seasonality

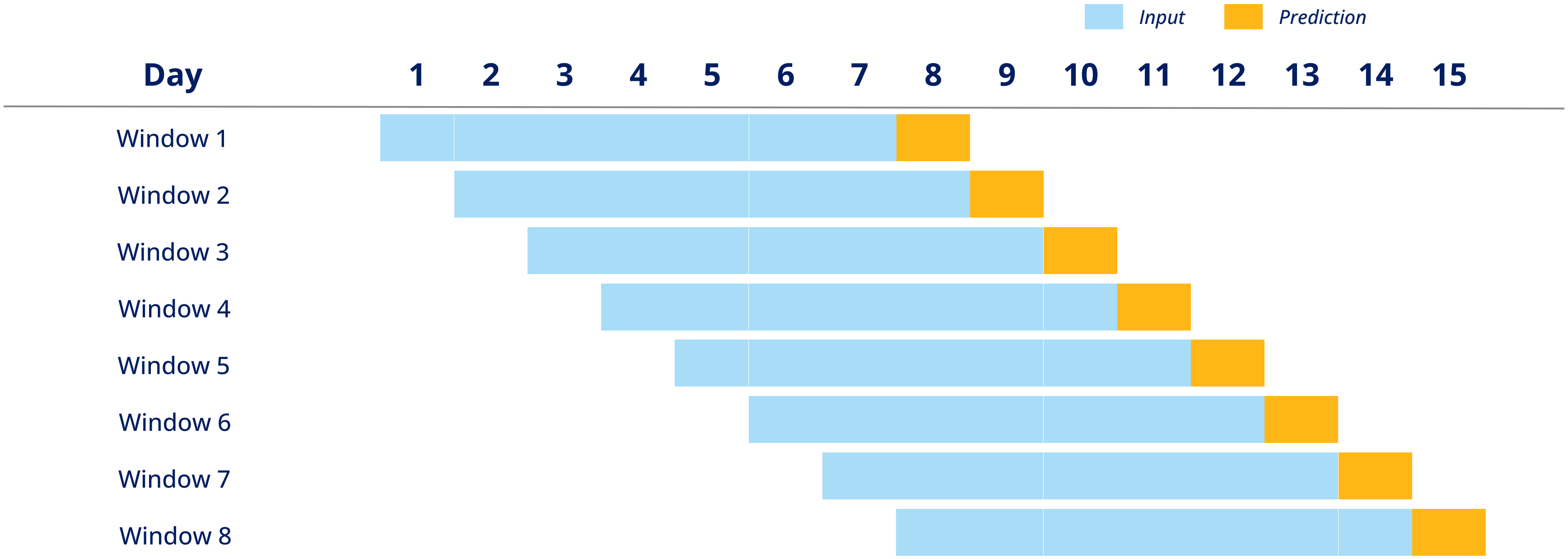
Events Feature

Adjusts for weekday / weekend / holiday surges

Price Dynamics

Understands promotional impact
(used only in LightGBM)

How the Model Learns to Forecast Demand



Our model learns from the **most recent 7 days** of sales to accurately forecast the next day's demand - repeating this process day by day to capture trends, events, and patterns

Not All Forecasting Mistakes Matter Equally...

Missing a fast-moving item by 10 units is more serious than missing a slow seller by the same amount

What We Measure

WRMSSE *Weighted Root Mean Squared Scale Error*

A forecasting score that puts more weight on:

- High-volume, high-value SKUs
- Lower levels in the hierarchy (store / item level)
- Long-term consistency (not just short-term fits)

Why It's Smart

Forecast Error Weight $\propto$ Product Importance

This means popular, high-volume SKUs and store-level forecasts count more

What That Means for You

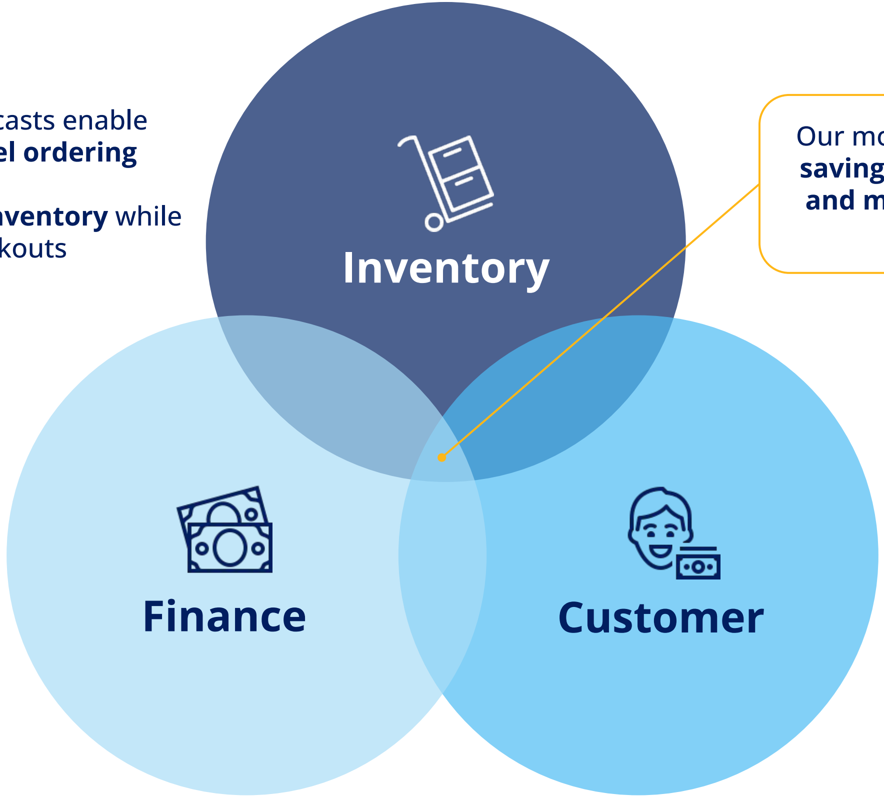
Our model performs best **where it matters most** - helping inventory managers prevent stockouts and make confident restocking decisions

Smarter Forecasts, Tangible Results

More accurate forecasts enable **smarter store-level ordering**

Helps maintain **lean inventory** while avoiding stockouts

Our model directly supports **cost savings, customer satisfaction, and more confident inventory decisions**



Reduces working capital tied up in **excess inventory**

Minimizes **markdowns and over-purchasing costs**

Improves **product availability**, especially during demand spikes

Enhances **customer satisfaction and retention**

Next Steps - Bringing the Model to the Store Floor

Forecast Dashboard Delivery

Integrate daily GRU-based forecasts into **existing inventory systems** (Power BI / Tableau)

Inventory managers can view **28-day demand predictions** per SKU, with confidence bands to guide restocking

Automated Replenishment Rules

Use forecast thresholds to **trigger reorder suggestions** or safety stock alerts

Create simple rules:
If predicted stockout in 7 days → suggest reorder today

Weekly Model Refresh Pipeline

Schedule **model runs** using Airflow or cloud jobs (eg., AWS Lambda / GCP Cloud Functions)

Automatically pull in n**ew sales + event data** each week and push updated forecasts to the dashboard

Thank You!

Zafrin Ahmed
Aakash Patil
Aditya Ravikrishnan
Mayank Singh
Archita Vaje